

# Algoritmos de Monte Carlo e Cadeias de Markov – CPS 767

## 2020/1

Prof. Daniel R. Figueiredo

### Terceira Lista de Exercícios

**Dica:** Para ajudar no processo de aprendizado responda às perguntas integralmente, mostrando o desenvolvimento das respostas.

#### Questão 1: Calculando $\sqrt{2}$

Vimos em aula um algoritmo de Monte Carlo para calcular o valor de  $\pi$  utilizando a relação entre áreas. Inspirado nesta mesma ideia, construa um algoritmo de Monte Carlo para calcular o valor de  $\sqrt{2}$ .

1. Descreva a variável aleatória cujo valor esperado está relacionado com  $\sqrt{2}$ . Obtenha analiticamente o valor esperado da sua variável aleatória.
2. Calcule analiticamente a variância dessa variável aleatória.
3. Implemente o método de Monte Carlo para gerar amostras da sua variável aleatória, calculando a média amostral  $M_n$  e utilizando-a para estimar  $\sqrt{2}$ . Sua função deve usar o algoritmo de Mersenne Twister para geração de números pseudo-aleatórios uniformes entre 0 e 1 (usar este algoritmo em todos os problemas da lista).
4. Seja  $\hat{e}_n$  o valor do estimador após  $n$  amostras. Trace um gráfico do erro relativo do estimador, ou seja  $|\hat{e}_n - \sqrt{2}|/\sqrt{2}$  em função de  $n$ , para  $n = 1, \dots, 10^6$  (utilize escala log – log no gráfico). O que você pode concluir?

#### Questão 2: Transformada inversa

Utilize o método da transformada inversa para gerar amostras de uma v.a.  $X$  com as seguintes densidades:

1. Distribuição exponencial com parâmetro  $\lambda > 0$ , cuja função densidade é dada por  $f_X(x) = \lambda e^{-\lambda x}$ , para  $x \geq 0$ .
2. Distribuição de Pareto com parâmetros  $x_0 > 0$  e  $\alpha > 0$ , cuja função densidade é dada por  $f_X(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}$ , para  $x \geq x_0$ .

#### Questão 3: Contando domínios na Web

Quantos domínios web existem dentro da UFRJ? Mais precisamente, quantos domínios existem dentro do padrão de nomes  $http://www.[a-z](k).ufrj.br$ , onde  $[a-z](k)$  é qualquer sequência de caracteres de comprimento  $k$  ou menor? Construa um algoritmo de Monte Carlo para estimar este número.

1. Descreva a variável aleatória cujo valor esperado está relacionado com a medida de interesse. Obtenha analiticamente o valor esperado da sua variável aleatória.
2. Calcule a variância dessa variável aleatória.
3. Implemente o método de Monte Carlo para gerar amostras da sua variável aleatória. Ou seja, você deve consultar o domínio gerado para determinar se o mesmo existe (utilize uma biblioteca `www` para isto).
4. Assuma que  $k = 4$ . Seja  $\hat{w}_n$  o valor do estimador do número de domínios após  $n$  amostras. Trace um gráfico de  $\hat{w}_n$  em função de  $n$  para  $n = 1, \dots, 10^4$  (ou mais, se conseguir). O que você pode dizer sobre a convergência de  $\hat{w}_n$ ?

#### Questão 4: Gerando amostras Normais

Seja  $Z$  uma variável aleatória com distribuição Normal com média 0 e variância 1. Em particular, a função densidade de  $Z$  é dada por  $f_Z(x) = 1/(\sqrt{2\pi})e^{-x^2/2}$ , com  $-\infty < x < \infty$ . Repare que  $Z$  assume valores positivos e negativos, mas com caudas que possuem a mesma probabilidade. Ou seja,  $P[Z \geq z] = P[Z \leq -z]$ , para todo  $z \geq 0$ . Construa um gerador de números aleatórios para  $Z$ . Dica: Utilize o método de amostragem por rejeição!

#### Questão 5: Estimando somas com *Importance Sampling*

Considere o problema visto em aula, de aplicar o método de Monte Carlo para estimar o valor de  $G_N = \sum_{i=1}^N i \log(i)$ . Use sua intuição para encontrar uma função de probabilidade proponente,  $h(i)$ , que tenha variância inferior à distribuição uniforme, ou seja,  $h(i) = 1/N$ .

1. Assuma que  $N = 10^6$ . Calcule numericamente o segundo momento do estimador usando a  $h(i)$  proposta.
2. Mostre como gerar amostras para a distribuição  $h(i)$ . Dica: método da transformada inversa!
3. Implemente o método de Monte Carlo para estimar o valor de  $G_N$ . Trace um gráfico do erro relativo do estimador, em função de  $n = 1, \dots, 10^7$  (calcule o valor exato da soma para determinar o erro relativo).

#### Questão 6: Integração de Monte Carlo

Considere a função  $f(x) = x^\alpha$  com  $\alpha > 0$ . Defina  $g(\alpha, a, b) = \int_a^b f(x)dx$  com  $0 \leq a < b$ , como sendo a integral de  $f(x)$  no intervalo  $[a, b]$ . Iremos calcular  $g$  usando Monte Carlo.

1. Determine analiticamente o valor de  $g(\alpha, a, b)$ . Dica: lembre Cálculo I.
2. Descreva a variável aleatória cujo valor esperado está relacionado com  $g(\alpha, a, b)$ . Obtenha analiticamente o valor esperado da sua variável aleatória.
3. Implemente o método de Monte Carlo para gerar amostras da sua variável aleatória, calculando a média amostral  $M_n$  e utilizando-a para estimar  $g(\alpha, a, b)$ . Repare que  $\alpha, a, b$  são parâmetros do seu programa.
4. Seja  $\hat{g}_n$  o valor do estimador após  $n$  amostras. Trace um gráfico do erro relativo do estimador, ou seja  $|\hat{g}_n - g(\alpha, a, b)|/g(\alpha, a, b)$  em função de  $n$ , para  $n = 1, \dots, 10^6$  (utilize escala log – log no gráfico). Utilize os seguintes valores para os parâmetros:  $\alpha = \{1, 2, 3\}$ ,  $a = 0$ ,  $b = \{1, 2, 4\}$ . O que você pode concluir em relação ao erro e os parâmetros?

#### Questão 7: Gerando subconjuntos

Considere  $S_{k,n}$  um espaço amostral dado por todos os subconjunto de tamanho  $k$  dentre  $n$  objetos. Assuma que cada elemento deste espaço amostral tem a mesma probabilidade, dada por  $1/|S_{k,n}|$ .

1. Descreva um algoritmo eficiente para gerar amostras deste espaço. Dica: pense em permutação!
2. Implemente seu algoritmo. Calcule o tempo médio de execução fazendo  $r = 10^3$  rodadas independentes para os seguintes parâmetros  $n = \{10^4, 10^6, 10^8\}$ ,  $k = \{10^1, 10^2, 10^3, 10^4\}$ . Trace um gráfico com valores de  $k$  no eixo-x e diferentes curvas com os valores de  $n$ . O que você pode concluir?